



42
JCS84 U.S. PTO
09/343929
06/30/99

Bescheinigung

Die Firma International Business Machines Corporation in Armonk, N.Y./V.St.A. hat eine Patentanmeldung unter der Bezeichnung

"Verfahren zur automatischen Generierung einer textlichen Äußerung aus einer Bedeutungsrepräsentation durch ein Computersystem"

am 9. Oktober 1998 beim Deutschen Patent- und Markenamt eingereicht.

Die beigehefteten Stücke sind eine richtige und genaue Wiedergabe der ursprünglichen Unterlagen dieser Patentanmeldung.

Die Anmeldung hat im Deutschen Patent- und Markenamt vorläufig das Symbol G 06 F 17/20 der Internationalen Patentklassifikation erhalten.

München, den 24. März 1999

Deutsches Patent- und Markenamt

Der Präsident

Im Auftrag

Hoi.

Aktenzeichen: 198 49 855.1

CERTIFIED COPY OF
PRIORITY DOCUMENT

THIS PAGE BLANK (USPTO)

B E S C H R E I B U N G

Verfahren zur automatischen Generierung einer textlichen
Äußerung aus einer Bedeutungsrepräsentation durch ein
Computersystem

Die Erfindung betrifft ein Verfahren zur automatischen Generierung einer textlichen Äußerung aus einer Bedeutungsrepräsentation durch ein Computersystem.

Bei einer immer größer werdenden Anzahl von Anwendungen eines Computersystems ist es vorteilhaft oder gar notwendig, bestimmte Inhalte an den Benutzer zu kommunizieren. Beispielsweise bei einem Fahrplanauskunftssystem sollte das Computersystem die erwünschte Fahrplanauskunft als Text dem Benutzer anzeigen können. Der Text sollte dabei nicht nur z.B. die Abfahrzeiten enthalten, sondern sollte dem Benutzer alle wesentlichen Informationen in der Form von vollständigen Sätzen oder sonstigen textlichen Äußerungen ausgeben. Diese Informationen können schriftlich auf einem Bildschirm angezeigt oder akustisch durch eine entsprechende natürlichsprachliche Ausgabe über einen Lautsprecher ausgegeben werden.

Die Erzeugung von Text, insbesondere die Generierung von Sätzen durch ein Computersystem ist bekannt.

So ist es möglich, daß in einem Speicher des Computersystems bestimmte Sätze vollständig abgespeichert werden, um von den einzelnen Programmen wieder aufgerufen und zur Anzeige gebracht zu werden. Beispielsweise kann der Satz "Datei wurde nicht gefunden." abgespeichert und zu gegebener Zeit wieder aufgerufen werden. Derartige Verfahren zur Generierung von Sätzen werden auch als "canned text systems" bezeichnet. Ersichtlich ist dieses Verfahren nicht für eine

große Anzahl von Sätzen und auch nicht für eine variable Satzgenerierung einsetzbar.

Ebenfalls ist es möglich, daß in dem Speicher nur Teile eines bestimmten Satzes abgespeichert sind, die bei deren Abruf von dem Computersystem variabel ergänzt werden. Beispielsweise kann der Satz "Die gewünschte Telefonnummer lautet ..." im Speicher des Computersystems abgespeichert sein. Bei dessen Abruf kann das Computersystem die betreffende Telefonnummer einsetzen und den vollständigen Satz an den Benutzer ausgeben. Ein solches System zur Satzgenerierung wird auch als "template system" bezeichnet. Es erlaubt zwar eine gewisse variable Satzgenerierung, ist aber in seiner Flexibilität trotzdem noch sehr begrenzt.

Schließlich ist es möglich, mit Hilfe von wissensbasierten linguistischen Systemen verschiedenartige Sätze durch ein Computersystem variabel zu erzeugen. Dabei werden von dem Computersystem sogenannte Bedeutungsrepräsentationen ermittelt, die beispielsweise die an den Benutzer auszugebenden Informationen in einem nicht-satzgebundenen und nicht-grammatikalischen Format enthalten. Diese muß von dem Computersystem in textliche Äußerungen oder vollständige Sätze umgewandelt werden. Derartige Systeme zur Generierung von textlichen Äußerungen aus Bedeutungsrepräsentationen erfordern einen hohen Aufwand an Speicherplatz, Rechenzeit und Entwicklungstätigkeit, so daß sie bisher nur für bestimmte, eng begrenzte Generierungsaufgaben vorhanden sind.

Aufgabe der Erfindung ist es, ein Verfahren zur automatischen Generierung einer textlichen Äußerung aus einer Bedeutungsrepräsentation durch ein Computersystem zur Verfügung zu stellen, das für möglichst große Anzahlen von Bedeutungsrepräsentationen und möglichst variabel eingesetzt werden kann, das aber trotzdem nur einen vertretbaren Aufwand an Speicherplatz und Rechenzeit erfordert.

Diese Aufgabe wird bei der Erfindung durch ein Verfahren zur automatischen Generierung einer Äußerung aus einer Bedeutungsrepräsentation nach dem Anspruch 1 gelöst.

Die Erfindung verwendet ein statistisches Modell zur Ermittlung einer Äußerung aus einer Bedeutungsrepräsentation. Bei der ermittelten Äußerung handelt es sich damit um die wahrscheinlichste Äußerung, die auf der Grundlage des statistischen Modells der vorgegebenen Bedeutungsrepräsentation zugeordnet werden kann.

Damit wird erreicht, daß die Vorteile der bekannten "canned text systems", insbesondere deren schnelle Zugreifbarkeit, weitgehend erhalten bleiben. Gleichzeitig wird jedoch die Anzahl der zur Verfügung stehenden Äußerungen nahezu unbegrenzt.

Die Entwicklung von Programmcode zu den Generierungsschritten wird von dem statistischen Modell und dessen Entwicklung streng getrennt. Damit wird die Erweiterung des Verfahrens und dessen gegebenenfalls erforderliche Anpassung wesentlich erleichtert.

Ein Wechsel beispielsweise von der deutschen zu der englischen Sprache kann allein durch einen Austausch der deutschsprachigen durch eine englischsprachige Übersetzung innerhalb des statistischen Modells erreicht werden, was eine erheblich Verminderung des Entwicklungsaufwands bedeutet.

Für die Ermittlung des statistischen Modells können bekannte, vorhandene Verfahren oder Programme verwendet werden, was ebenfalls zu einem geringeren Aufwand bei der Entwicklung des gesamten Verfahrens führt.

Ebenfalls ist es möglich, daß für nicht in dem statistischen Modell enthaltene Bedeutungsrepräsentationen zumindest

brauchbare textliche Äußerungen erzeugt werden. Damit ist das Verfahren auch zur Verarbeitung von neuen Bedeutungsrepräsentationen geeignet.

Des weiteren weist das erfindungsgemäße Verfahren keinen besonders hohen Speicherplatzbedarf auf und ist auch im Hinblick auf seine Rechenzeit bei der Generierung unkritisch. Es besitzt somit nicht die Nachteile von wissensbasierten linguistischen Systemen.

Durch die Überführung der Bedeutungsrepräsentationen und der zugehörigen Äußerungen in ein internes Format ist es möglich, den Korpus und das daraus entstehende statistische Modell zu vereinfachen. Insbesondere der Speicherplatzbedarf wird dadurch weiter vermindert. Ebenfalls können für die statistische Verarbeitung irrelevante Informationen ausgefiltert werden. Das Modell wird dadurch besser und kann schneller berechnet werden.

Durch die Einführung von Generalisierungen ist es möglich, die Anzahl der zur Verfügung stehenden Äußerungen praktisch unbegrenzt zu erweitern, ohne hierdurch die Anforderungen an den Speicherplatzbedarf und die Rechenzeit wesentlich zu erhöhen. Umgekehrt wird ein Korpus durch Zusammenfassung äquivalenter Einträge kompakter und das zu berechnende statistische Modell robuster und akkurater.

Weitere Merkmale, Anwendungsmöglichkeiten und Vorteile der Erfindung ergeben sich aus der nachfolgenden Beschreibung von Ausführungsbeispielen der Erfindung, die in den Figuren der Zeichnung dargestellt sind. Dabei bilden alle beschriebenen oder dargestellten Merkmale für sich oder in beliebiger Kombination den Gegenstand der Erfindung, unabhängig von ihrer Zusammenfassung in den Patentansprüchen oder deren Rückbeziehung sowie unabhängig von ihrer Formulierung bzw. Darstellung in der Beschreibung bzw. in der Zeichnung.

- Figur 1 zeigt eine schematische Darstellung eines Verfahrens zur Ermittlung eines statistischen Modells,
- Figur 2 zeigt eine schematische Darstellung eines Verfahrens zur Generierung von Äußerungen auf der Grundlage des statistischen Modells der Figur 1,
- Figur 3 zeigt eine schematische Darstellung eines Verfahrens zur Generalisierung eines Korpus, aus dem ein generalisiertes statistisches Modell entsprechend der Figur 1 ermittelt wird, und
- Figur 4 zeigt eine schematische Darstellung eines Verfahrens zur Generierung von Äußerungen auf der Grundlage des generalisierten statistischen Modells der Figur 3.

Ausgangspunkt für das Verfahren zur Berechnung eines statistischen Modells nach der Figur 1, wie auch für das Verfahren zur Generierung von Äußerungen nach der Figur 2 sind Paare von Bedeutungsrepräsentationen und zugehörigen Äußerungen. Aus einer Vielzahl derartiger Paare wird in einem vorbereitenden Schritt - dem Verfahren nach der Figur 1 - ein statistisches Modell ermittelt, das danach - im Verfahren nach der Figur 2 - immer wieder zum Zwecke der Generierung von Sätzen durchsucht wird.

Die Darstellungsweise von Bedeutungsrepräsentationen und zugehörigen Äußerungen ist vorgegeben und wird als extern bezeichnet. Diese externe Darstellungsweise der Paare von Bedeutungsrepräsentationen und zugehörigen Äußerungen kann beispielsweise wie folgt gegeben sein:

Bedeutungsrepräsentation, externe Darstellungsweise:

```
from_until(15,14),  
def(17,i2,11,ht190),  
mofy(13,i1,apr),  
dofm(14,i2,i1,27),  
dofm(15,i2,i3,23) ...
```

zugehörige Äußerung, externe Darstellungsweise:

"From the 23rd to the 27th of April"

Es ist möglich, daß diese externe Darstellungsweise unabhängig von anderen Programmen des Computersystems erstellt wird. Ebenfalls ist es möglich, daß sich diese oder eine beliebige andere Darstellungsweise aus der Verwendung von anderen Programmen des Computersystems ergibt.

Beispielsweise kann sich die Darstellungsweise aus der Verwendung eines Übersetzungsprogramms oder eines Spracherkennungsprogramms oder dergleichen ergeben, die die jeweilige externe Darstellungsweise bereits verwenden.

In der externen Darstellungsweise werden eine Vielzahl von Paaren von Bedeutungsrepräsentationen und von zugehörigen Äußerungen in das Computersystem eingegeben und dort in einem Speicher abgespeichert.

Die Paare von Bedeutungsrepräsentationen und zugehörigen Äußerungen können entweder manuell dem Computersystem eingegeben werden. In diesem Fall kann eine Person genau diejenigen Bedeutungsrepräsentationen und zugehörigen Äußerungen in das Computersystem eingeben, die für die betreffende Anwendung sinnvoll oder notwendig sind. Ebenfalls ist es möglich, daß die Paare von Bedeutungsrepräsentationen und zugehörigen Äußerungen von einem anderen Programm, beispielsweise von dem

Übersetzungsprogramm oder dem Spracherkennungsprogramm übernommen und in dem Computersystem abgespeichert werden.

Sämtliche in das Computersystem eingegebene Paare von Bedeutungsrepräsentationen und zugehörigen Äußerungen stellen einen Korpus dar. Je umfangreicher dieser Korpus ist, desto umfangreicher sind auch die Möglichkeiten einer späteren Generierung von Äußerungen durch das Computersystem.

Ausgehend von diesem Korpus ermittelt das Computersystem nach dem Verfahren der Figur 1 ein statistisches Modell.

Hierzu wird der Korpus in einem Schritt 11 zuerst von der externen Darstellungsweise in ein internes Format überführt. Die Paare C werden also in Paare C_{internal} abgebildet. Zur Durchführung dieses Schritts wird eine Negativliste 12 und eine Übersetzungstabelle 13 verwendet. Das interne Format stellt eine Vereinfachung der externen Darstellungsweise des Korpus dar, die auf die durchzuführende Ermittlung eines statistischen Modells ausgerichtet bzw. auf diese beschränkt ist.

Die Negativliste beschreibt Informationen, die gegebenenfalls in der externen Darstellung enthalten sind, die aber für das statistische Modell irrelevant oder sogar schädlich sind. Diese Informationen finden sich in der internen Darstellung nicht wieder. Die Übersetzungstabelle stellt den Bezug von Teilen der externen Darstellung zu Teilen der internen Darstellung her.

Beispielsweise wird das eingangs genannte Paar einer Bedeutungsrepräsentation und der zugehörigen Äußerung wie folgt von der externen Darstellungsweise in das interne Format überführt:

Bedeutungsrepräsentation, internes Format:

```
from_until,def,i1,i2,i3,i1,...,i7,  
dofm(27th,23rd),mofy(April)
```

zugehörige Äußerung, internes Format:

```
from,the,23rd,to,the,27th,of,April
```

Der nunmehr im internen Format vorhandene Korpus bildet die Grundlage zur Ermittlung eines statistischen Modells. Dazu werden von dem Computersystem in einem Schritt 14 eine Vielzahl von Fragen Qstart bestimmt, mit denen die einzelnen Einträge des Korpus klassifiziert werden können. Darüber hinaus ist es möglich, daß auch manuell eine Anzahl von initialisierenden Fragen 15 vorgegeben werden können.

Aus diesen Fragen berechnet das Computersystem anhand von bekannten Verfahren in einem Schritt 16 zu dem vorhandenen Korpus ein statistisches Modell. Dabei werden zu jedem Bestandteil der vorhandenen Bedeutungsrepräsentationen von dem Computersystem Fragen erzeugt, so daß die jeweiligen Bestandteile klassifiziert werden. Auf diese Weise ermittelt das Computersystem beispielsweise ein statistisches Entscheidungsbaummodell T.

Im Hinblick auf die Erzeugung dieser Fragen können vorgegebene Randbedingungen eine Rolle spielen, beispielsweise daß das entstehende Entscheidungsbaummodell möglichst ausgewogen sein soll oder dergleichen.

Das im Schritt 16 berechnete Modell T wird im Computersystem abgespeichert. Die Ermittlung dieses Modells T, also das Verfahren nach der Figur 1 wird einmalig durchlaufen, und zwar vor der eigentlichen Generierung von Ausdrücken nach dem Verfahren der Figur 2. Das letztgenannte Verfahren erfolgt danach auf der Grundlage des ermittelten Modells T.

Bei dem Verfahren nach der Figur 1 werden also Paare C von Bedeutungsrepräsentationen und deren zugehörige Äußerungen von der externen Darstellungsweise in das interne Format überführt, um danach aus den Paaren C_{internal} das statistische Modell T zu ermitteln. Diese Schritte werden automatisch von dem Computersystem durchgeführt.

Auf der Grundlage des statistischen Modells T kann das Computersystem nach dem Verfahren der Figur 2 Äußerungen generieren.

Hierzu wird dem Computersystem eine Bedeutungsrepräsentation als Eingabe i zugeführt. Diese wird in einem Schritt 21 zuerst wieder von der externen Darstellungsweise in das interne Format überführt. Hierzu werden wieder die Negativliste 12 und die Übersetzungstabelle 13 verwendet. Nach dem Schritt 21 liegt somit die Bedeutungsrepräsentation als Eingabe i_{internal} vor.

In einem Schritt 22 wird das statistische Modell T durchsucht. Dieses ist beispielsweise - wie bereits erläutert wurde - als Entscheidungsbaummodell 23 in dem Computersystem abgespeichert. Es werden somit zu der in das interne Format überführten Bedeutungsrepräsentation die bereits erwähnten Fragen gestellt und es werden beispielsweise parallel eine bestimmte Anzahl alternativer Antworten weiterverfolgt, die aufgrund des Modells T und der Eingabe i die höchste Wahrscheinlichkeit besitzen.

Auf diese Weise wird eine statistische Suche in dem Entscheidungsbaummodell von dem Computersystem durchgeführt, mit der die am besten passende Äußerung zu der vorgegebenen Bedeutungsrepräsentation ermittelt wird. Diese Äußerung stellt eine Ausgabe o_{internal} dar, die noch in dem internen Format vorliegt.

In einem Schritt 24 wird das interne Format der Ausgabe ointernal wieder in die externe Darstellungsweise rücküberführt, so daß nunmehr die ermittelte Äußerung als Ausgabe o in der externen Darstellungsweise vorliegt.

Bei dem Verfahren nach der Figur 2 wird also die Eingabe i in das interne Format überführt. Dann wird die Eingabe iinternal anhand des statistischen Modells klassifiziert, um die am besten passende Ausgabe ointernal zu ermitteln. Diese wird schließlich wieder in die Ausgabe o der externen Darstellungsweise rücküberführt. Diese Schritte werden automatisch von dem Computersystem durchgeführt.

Gegebenenfalls ist es möglich, daß nach der Ermittlung der am besten passenden Äußerung noch die Korrektheit dieser Äußerung mit Hilfe von zusätzlichen Verfahren durch das Computersystem überprüft wird.

Das Verfahren nach der Figur 2 und gegebenenfalls die genannten zusätzlichen Verfahren werden für jede Eingabe i durchlaufen, so daß für jede Eingabe i, also für jede Bedeutungsrepräsentation, eine zugehörige Ausgabe o, also eine zugehörige Äußerung von dem Computersystem erzeugt wird.

Zur Erhöhung der Leistungsfähigkeit der vorbeschriebenen Verfahren ist es möglich, den aus Paaren von Bedeutungsrepräsentationen und zugehörigen Äußerungen bestehenden Korpus zu generalisieren.

Zu diesem Zweck werden bestimmte, miteinander korrespondierende Teile der Bedeutungsrepräsentationen und der zugehörigen Äußerungen, die sich verallgemeinern lassen, und die immer wieder auftreten, durch sogenannte Patterns bzw. Templates verallgemeinert. So wird beispielsweise ein bestimmter Monat nicht mehr als solcher in den Korpus übernommen, sondern als verallgemeinerter Monat. Bei der

Generierung von Äußerungen wird dann die Zuordnung des verallgemeinerten Monats zu dem bestimmten Monat vorab in dem Computersystem abgespeichert, um danach in die ermittelte Äußerung wieder eingesetzt werden zu können.

Beispielhaft kann eine Generalisierung der Bedeutungsrepräsentationen und der zugehörigen Äußerungen wie folgt von dem Computersystem vorgenommen werden:

Bedeutungsrepräsentationen: zugehörige Äußerungen:

externe Darstellungsweise:

mofy(13,i1,apr),
dofm(14,i2,i1,27),
dofm(16,i2,i3,23),

"From the 23rd to
the 27th of April"

internes Format:

from_until,def,i1,i2,i3,
11,...,17,
dofm(27th,23rd),
mofy(April)

from,the,23rd,to,
the,27th,of,April

generalisiertes Format:

from_until,def,i1,i2,i3,
11,...,17,
dofm(Vdofm1,Vdofm2)
mofy(Vmofy1)

from,the,Vdofm2,to
the,Vdofm2,of,Vmofy1

In dem generalisierten Format ist beispielsweise der bestimmte Monat "April" durch den verallgemeinerten Monat "Vmofy1" ersetzt. Der Korpus enthält also nicht mehr den bestimmten Monat "April", sondern nur noch den verallgemeinerten Monat "Vmofy1". Entsprechendes gilt z.B. für den bestimmten Monat "Juni", der in dem Korpus ebenfalls

durch denselben verallgemeinerten Monat "Vmofyl" repräsentiert wird. Durch die Generalisierung werden also in diesem Fall die bestimmten Monate des Jahres "Januar, Februar, ... Dezember" durch den verallgemeinerten Monat "Vmofyl" substituiert.

Ausgehend von den im internen Format vorliegenden Paaren Cinternal von Bedeutungsrepräsentationen und zugehörigen Äußerungen ermittelt das Computersystem entsprechend der Figur 3 Paare Cgeneralized, die dann in einem generalisierten Format vorliegen. Diese Generalisierung wird für alle einzelnen Paare cinternal durchgeführt, die jeweils in Paare cgeneralized überführt werden.

Bei dem Verfahren nach der Figur 3 wird jedes Paar des Korpus von dem Computersystem auf mögliche Substitutionen überprüft. Die Substitutionen können dabei manuell vorgegeben werden, oder sie können gegebenenfalls auch von dem Computersystem automatisch ermittelt werden. Die Substitutionen sind in dem Speicher des Computersystems abgespeichert. In der Figur 3 sind die Substitutionen mit der Bezugsziffer 31 gekennzeichnet.

Auf der Grundlage dieser Substitutionen 31 wird die in dem internen Format vorliegende Bedeutungsrepräsentation iinternal eines bestimmten Paares cinternal in einem Schritt 32 in eine in dem generalisierten Format vorliegende Bedeutungsrepräsentation ipattern überführt. In entsprechender Weise wird die in dem internen Format vorliegende zugehörige Äußerung ointernal des bestimmten Paares cinternal in einem Schritt 33 in eine in dem generalisierten Format vorliegende Äußerung otemplate überführt. Insgesamt entsteht damit zu jedem in dem internen Format vorliegenden Paar cinternal ein in dem generalisierten Format vorliegendes Paar cgeneralized, das sich aus der jeweiligen Bedeutungsrepräsentation ipattern

und der zugehörigen Äußerung otemplate jeweils im generalisierten Format zusammensetzt.

Danach wird aus den Paaren Cgeneralized in derselben Weise ein generalisiertes statistisches Modell T' durch das Computersystem ermittelt, wie dies im Zusammenhang mit dem Verfahren nach der Figur 1 bereits erläutert worden ist. Das entstehende Entscheidungsbaummodell unterscheidet sich von demjenigen der Figur 1 im wesentlichen dadurch, daß in ihm die abgespeicherten Substitutionen 31 enthalten sind, daß also anstelle von bestimmten Monaten "April" oder "Juni" oder dergleichen deren Substitution "Vmofyl" vorhanden ist.

Auf der Grundlage des generalisierten statistischen Modells T' kann das Computersystem nach dem Verfahren der Figur 4 Äußerungen generieren.

Hierzu wird dem Computersystem eine Bedeutungsrepräsentation als Eingabe i zugeführt. Diese wird in einem Schritt 41 zuerst wieder von der externen Darstellungsweise in das interne Format überführt. Hierzu werden wieder die Negativliste 12 und die Übersetzungstabelle 13 verwendet. Nach dem Schritt 41 liegt somit die Bedeutungsrepräsentation als Eingabe iinternal vor.

In einem Schritt 42 wird das generalisierte statistische Modell T' durchsucht. Dieses ist als Entscheidungsbaummodell 43 in dem Computersystem abgespeichert. Auf diese Weise wird eine statistische Suche in dem Entscheidungsbaummodell von dem Computersystem durchgeführt, mit der die am besten passende Äußerung zu der vorgegebenen Bedeutungsrepräsentation ermittelt wird. Diese Äußerung kann - aufgrund der dem statistischen Modell T' zugrundeliegenden Generalisierungen - Substitutionen enthalten.

In einem Schritt 44 speichert das Computersystem etwa zeitgleich mit dem Schritt 42 die in den

Bedeutungsrepräsentationen iinternal tatsächlich vorkommenden Substitutionen mit ihrem jeweiligen bestimmten Inhalt in seinem Speicher ab.

Die sich aus dem Schritt 43 ergebenden vorläufigen Äußerungen und die in dem Schritt 44 gespeicherten Substitutionen werden in einem Schritt 45 derart überarbeitet, daß die vorhandenen Substitutionen wieder ersetzt werden. Es wird also der gespeicherte Inhalt nunmehr wieder anstelle der Substitution in die vorläufige Äußerung eingesetzt, so daß die endgültige Äußerung zu der eingangs vorgegebenen Bedeutungsrepräsentation entsteht. Diese Äußerung stellt eine Ausgabe ofilled dar, die noch in dem internen Format vorliegt.

In einem Schritt 46 wird das interne Format der Ausgabe ofilled wieder in die externe Darstellungsweise rücküberführt, so daß nunmehr die ermittelte Äußerung als Ausgabe o in der externen Darstellungsweise vorliegt.

P a t e n t a n s p r ü c h e

1. Verfahren zur automatischen Generierung einer textlichen Äußerung aus einer Bedeutungsrepräsentation durch ein Computersystem mit den folgenden Schritten:
 - a) es wird ein statistisches Modell auf einer Vielzahl von vorgegebenen Paaren von Bedeutungsrepräsentationen und zugehörigen Äußerungen durch das Computersystem ermittelt und abgespeichert,
 - b) es wird dem Computersystem eine Bedeutungsrepräsentation vorgegeben,
 - c) aus der Bedeutungsrepräsentation wird von dem Computersystem mittels des statistischen Modells eine zugehörige Äußerung ermittelt,
 - d) die Schritte b) und c) werden von dem Computersystem gegebenenfalls für weitere Bedeutungsrepräsentationen wiederholt.
2. Verfahren nach einem der vorstehenden Ansprüche, mit den weiteren Schritten:

vor dem Schritt a) werden die Paare von Bedeutungsrepräsentationen und zugehörigen Äußerungen in ein internes Format überführt,

nach dem Schritt b) wird die Bedeutungsrepräsentation in das interne Format überführt,

nach dem Schritt c) wird die ermittelte Äußerung aus dem internen Format rücküberführt.

3. Verfahren nach einem der vorstehenden Ansprüche, mit den weiteren Schritten:

vor dem Schritt a) werden die Paare von Bedeutungsrepräsentationen und zugehörigen Äußerungen mit Hilfe von Substitutionen in ein generalisiertes Format überführt,

nach dem Schritt c) wird die ermittelte Äußerung mit Hilfe der Substitutionen vervollständigt.

Z u s a m m e n f a s s u n g

Es wird ein Verfahren zur automatischen Generierung einer textlichen Äußerung aus einer Bedeutungsrepräsentation durch ein Computersystem beschrieben. Bei dem Verfahren wird ein statistisches Modell auf einer Vielzahl von vorgegebenen Paaren von Bedeutungsrepräsentationen und zugehörigen Äußerungen durch das Computersystem ermittelt und abgespeichert. Es wird dem Computersystem eine Bedeutungsrepräsentation vorgegeben, aus der von dem Computersystem mittels des statistischen Modells eine zugehörige Äußerung ermittelt wird. Diese Schritte werden von dem Computersystem gegebenenfalls für weitere Bedeutungsrepräsentationen wiederholt.

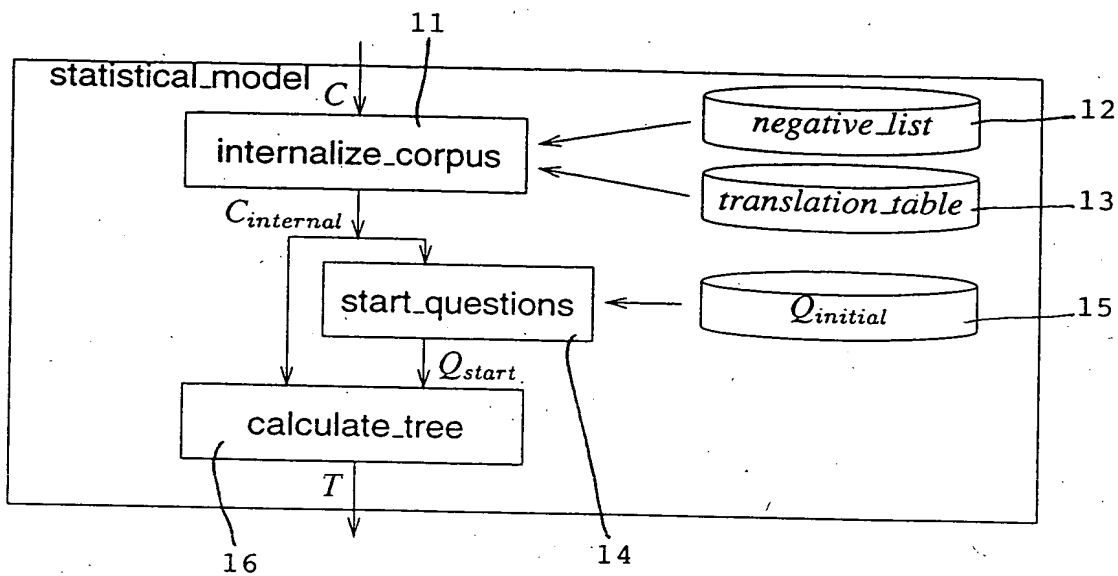


FIG. 1

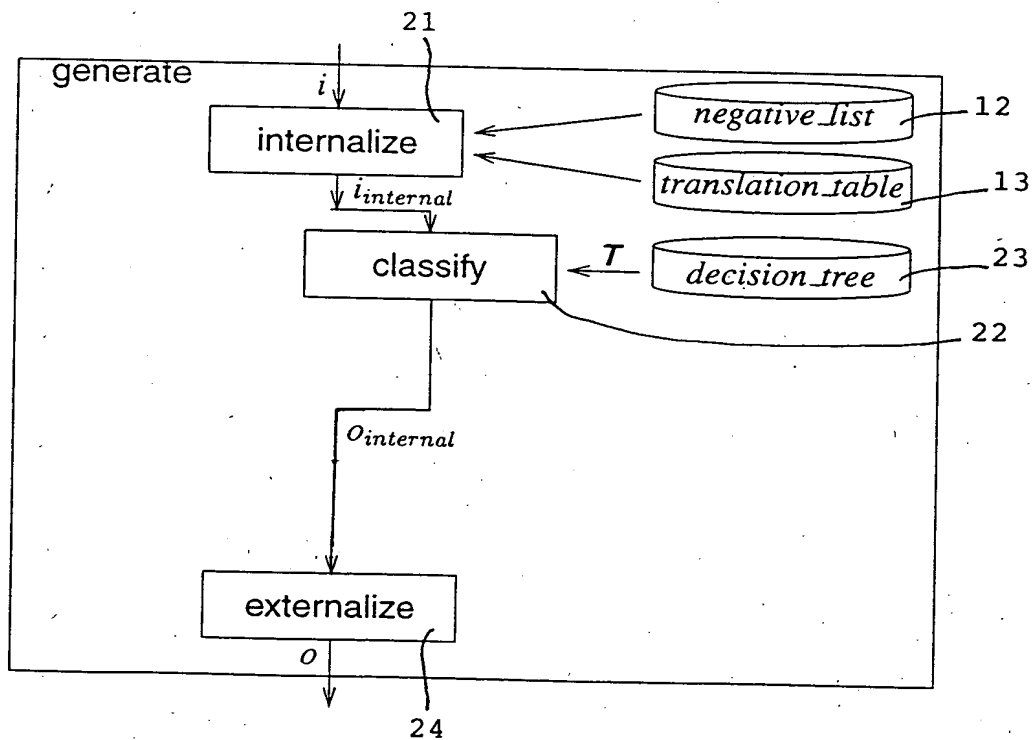


FIG. 2

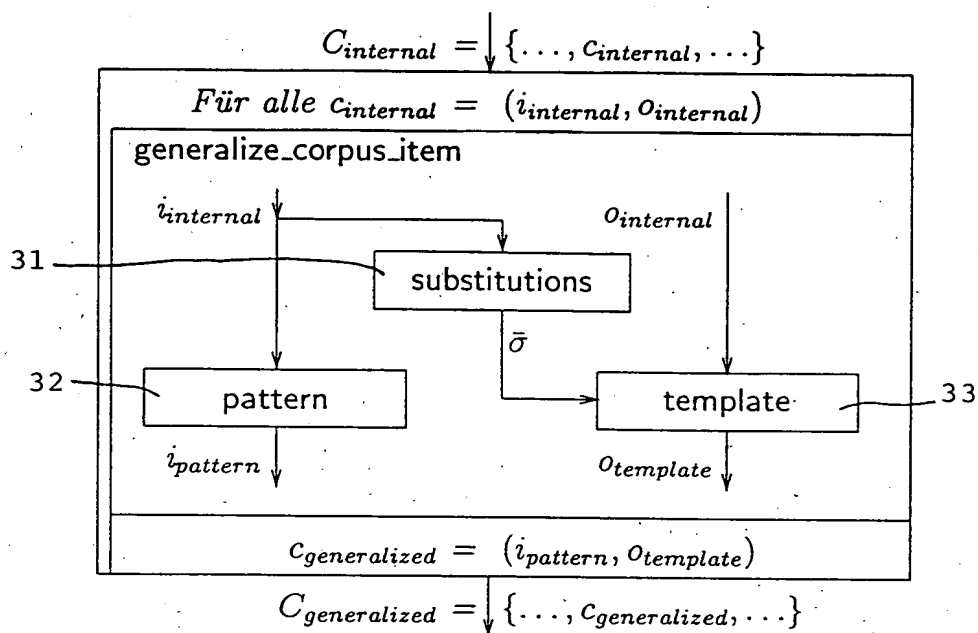


FIG. 3

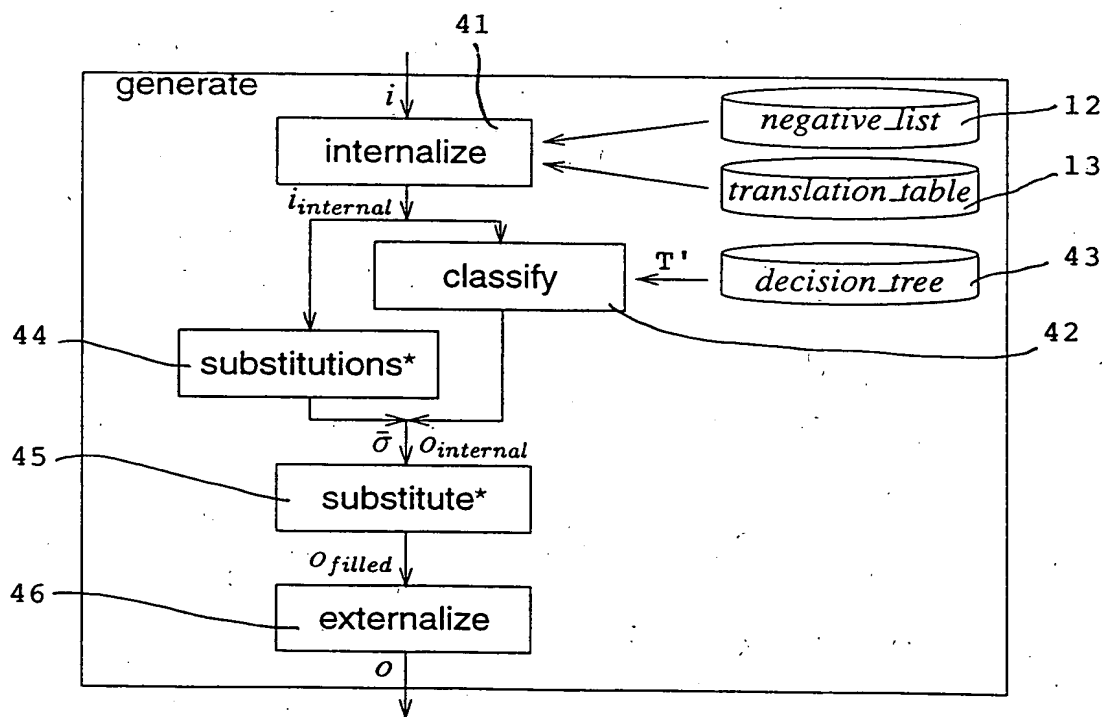


FIG. 4

THIS PAGE BLANK (USPTO)